

Counting with manifolds. An adaptively optimized algorithm for identifying cells in X-ray micro-CT scans of biomaterials.

A. Madra¹, A.Y. Lin¹, D.J. Vanselow¹, K.C. Cheng¹

¹ *Department of Pathology, College of Medicine, The Pennsylvania State University, {annamadra, ayl114, dxv46, kcheng}@psu.edu*

Résumé — Living organisms and, consequently, biomaterials are made of cells. Identifying them and characterizing their geometry and spatial distribution is the first step towards building multi-scale models of these materials. We propose a method to identify cells in an X-ray microtomographic scan of a zebrafish. To account for scanning artifacts and partial volume effect, the method is adaptively calibrated using parameters approximated from the manifold of manually selected and optimized special cases. The algorithm is validated for cells in the eyes of zebrafish specimens of different ages.

Mots clés — biomaterials, reduced-order modeling, X-ray microCT.

1 Introduction

Compared to engineering materials, biomaterials and living organisms are highly heterogeneous and form hierarchical structures at multiple length scales. Fortunately, as postulated by Theodor Schwann in [16], they all have a unit building block : a cell. Hence, to understand the changes in and mechanical behavior of biological materials, it is essential to identify cells, characterize their geometry, and observe their spatial distribution. The latter leads to the formation of organs – natural, highly heterogeneous composites.

In [1, 2], X-ray microtomography has been used to image the entire organism at a cellular resolution. This technique, referred to as X-ray histotomography, is unprecedented in the field of biology that mainly relies on 2.5D histology slides to provide information on the morphology of biomaterials. It also opened the possibility to create a first quantitative cellular model of an organism. The initial step towards creating this model consists of cell identification and counting.

Several methods can be used to identify cells. Most rely on the identification of the border of the cell, either by using gradient maps [10], watershed transform [13], or by altering specimen preparation protocol to enhance cell membranes [11] and thus simplify image processing. Recently, with the advances in deep learning, many approaches have been proposed that rely on Convolutional Neural Networks (CNN) and their derivatives [12, 13, 8, 10]. While these methods continue to improve in the accuracy of detection, even when applied to general cases, like in [10], they still require huge, manually annotated datasets to train or update the model. Such databases are becoming publicly available [7, 9], albeit their focus is on very-high-resolution microscopy, implying 2D data.

Overall, the existing approaches have two drawbacks

1. They rely on high-resolution data where pixel/voxel size is at least one tenth of the cell size, although one-hundredth is more common.
2. They are optimized for two-dimensional data sets.

The first drawback is critical in our case since the highest resolution for the large field-of-view X-ray microCT imaging still leads to cell size equal to 5-6 times the resolution. Hence, most automated segmentation methods like thresholding or methods involving gradient maps will not be efficient. Methods involving the watershed algorithm may be effective but are still prone to errors due to scan artifacts, and particularly to the prominent partial volume effect. Thus, they require local optimization to improve performance.

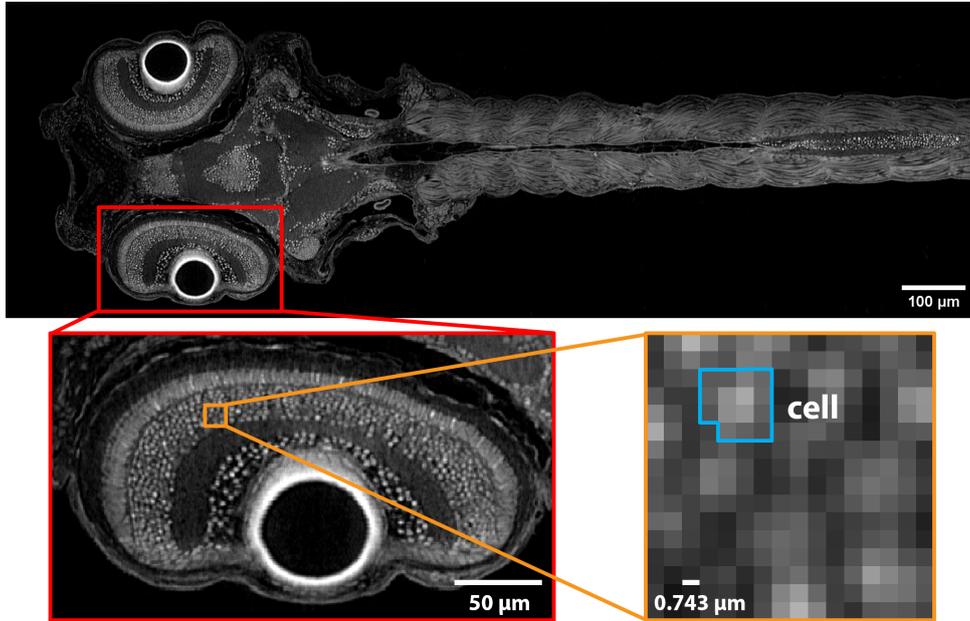


FIGURE 1 – A 2D slice of X-ray histotomography of a 5 dpf (days post fertilization) zebrafish (*Danio rerio*). Insets show the left eye and the cells of interest compared to voxel size.

The second drawback is of lesser importance computationally since most of the methods can be generalized into 3D. The challenge here is the difficulty of validation. While verification that cells are properly identified in 2D is relatively straightforward for both a non-expert and a pathologist or a biologist, it is not the case for 3D data. Even establishing a "ground truth" manually segmented data set is prone to variation depending on the operator's skill, usually requiring substantial workload and specialized equipment [15]. Limiting the number of cases for which the algorithms has to be manually calibrated is thus imperative.

In our method, we address both of these drawbacks by proposing a simple optimization algorithm to count the cells but calibrating it using a manifold of parameters obtained based on expert-validated special cases. This paper consists of the following : we begin by describing the cell counting algorithm, followed by the adaptive parameterization, and the results for counting cells in the left and right eye of a zebrafish (*Danio rerio*) at different stages of development. In conclusion, we state how this approach can be expanded to facilitate further characterization of biomaterials specimens.

2 Cell counting

The cells of interest are those in the eye of the zebrafish (Fig. 1). In the X-ray microCT scans, they are represented by groups of voxels with locally higher intensities, indicating higher X-ray attenuation. These groups are similar in shape to irregular ellipsoids. At the resolution in question : $0.743 \mu\text{m}$, they span 5-6 voxels in perpendicular directions. Since the goal is only to estimate the number of cells and not to segment the data set, the problem can be reduced to finding extrema in the 3D voxel intensity function within a neighborhood of size $X \times Y \times Z$ with a minimum distance D between results.

The solution to this problem is described in Alg. 1. First, we inspect the 3D subsets of size $X \times Y \times Z$ and return the maxima within each. Then, we remove the maxima that fall below the threshold T of minimum voxel intensity, and finally, the maxima that are within D distance of each other are averaged into one result.

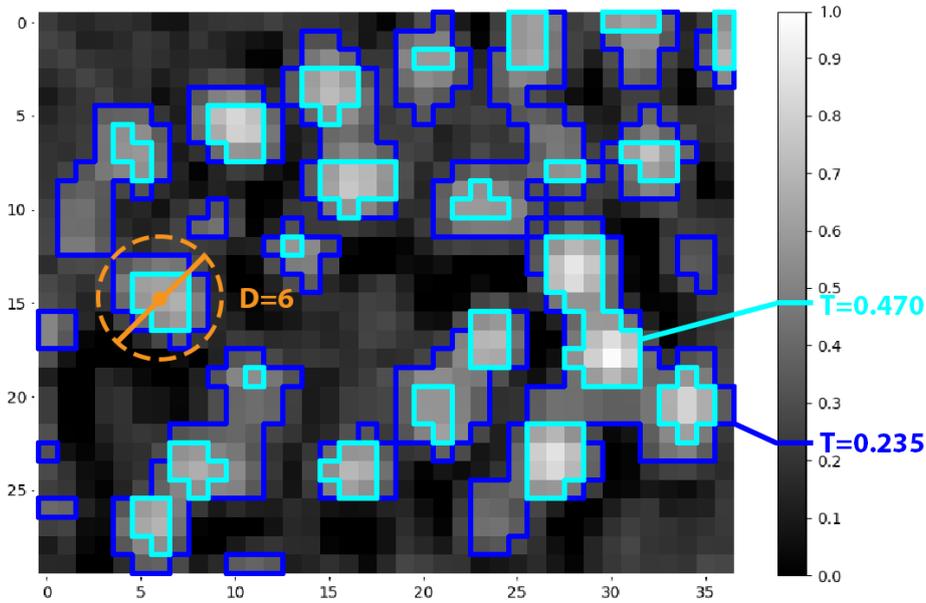


FIGURE 2 – Parameters of the algorithm : intensity threshold T and distance D .

Data: 3D array of voxel intensities, sizes X, Y, Z , distance D , threshold T

Result: number of cells N

```

for all voxel groups of size  $X \times Y \times Z$  do
  | find coordinates of voxel with max intensity;
end

for all found maxima do
  | if maximum  $> T$  and no maximum exists within  $D$  then
  |   | add 1 to  $N$ ;
  | else if maximum  $> T$  then
  |   | average maximum with neighbors;
  |   | remove neighbors from maxima list;
  |   | add 1 to  $N$ ;
  | else
  |   | remove from maxima list;
end

```

Algorithm 1: Cell counting.

3 Adaptive calibration with the manifold of parameters

The cell counting algorithm described in the previous section is straightforward, but its performance heavily relies on the choice of parameters X, Y, Z, D , and T . X-ray microCT data is subject to partial volume effect when the size of the object of interest is approaching the scan resolution. This means that objects imaged at a higher resolution consist of areas with homogeneous voxel intensity and have distinct borders, while at lower resolutions, the homogeneous area of voxel intensity is surrounded by a fringe of the average of the object's intensity and its surroundings. This implies that if the image in question contains only cells and homogeneous background, then the choice of intensity threshold T is equivalent to the designation of the border of a cell area (Fig. 2).

Thus we can reduce the parameters considered to T and D . While this is a manageable number to optimize for computations, it is nevertheless open to interpretation. For the dataset in question, manual adjustment resulted in a range of cell counts from 50 to 72 thousand cells, implying an almost 30 percent difference. While the choice of D is intuitive and can be easily computed given the maximum sizes of cells, the choice of T depends on the following :

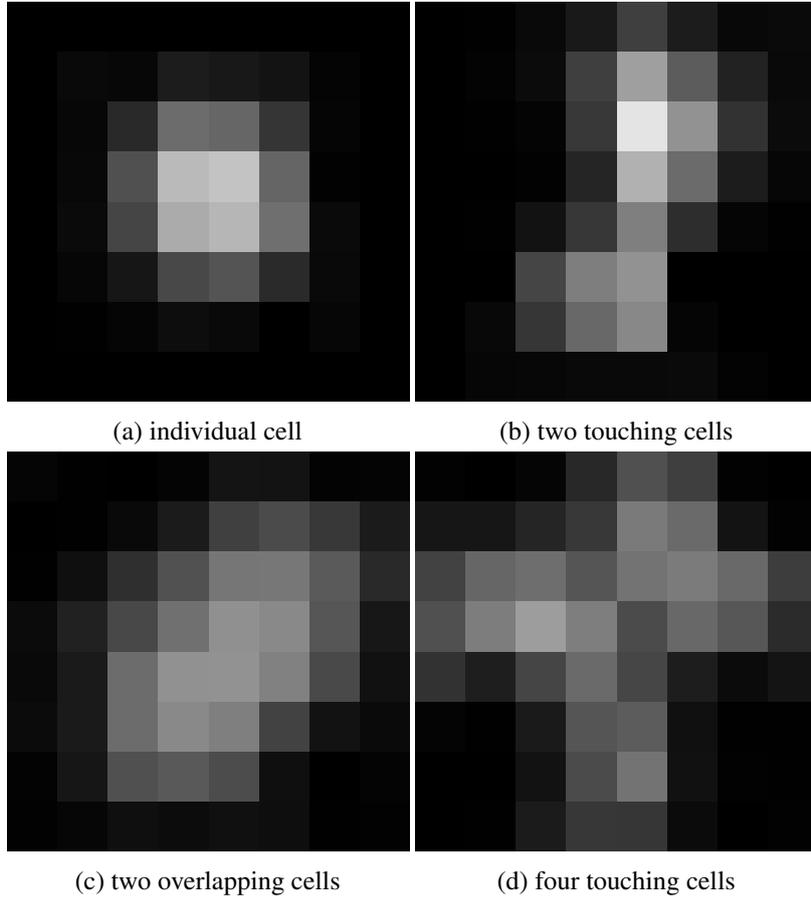


FIGURE 3 – Special cases : a) individual cell; b) two touching cells; c) two overlapping cells; d) four touching cells.

1. Intensity values in the scan, depending on the X-ray microCT setup, specimen geometry, and reconstruction algorithm.
2. Resolution of the scan and the prevalence of the partial-volume effect.
3. Spacing of the cells.

While the first two can be controlled and standardized, it is not yet practicable for exploratory X-ray microCT scans. The third varies throughout the specimen (Fig. 2) and in areas of high cell density can cause identification of false positives. In those areas, the intensity threshold T needs to be higher to prevent the merging of neighboring cells into one central cell.

To consider these special cases, we compiled a data set of representative voxel regions. Here, to ease comprehension, we show examples of their 2D equivalents in Fig. 3. The optimal values of T were then determined by a human expert for each special case.

The key part of the algorithm is determining which case is represented by the data and then looking up the T value. To measure the similarity of the analyzed 3D subset of the data and special cases, we propose to compare principal components of their snapshots s_m

$$s_m = \begin{bmatrix} \vdots \\ \mu(x, y, z) \\ \vdots \end{bmatrix} \in \mathbb{R}^N \quad (1)$$

where μ is the X-ray intensity of a voxel at a location x, y, z and $m = 1, \dots, M$ with M – number of expert-optimized special cases. The snapshots are collected in a matrix S

$$S = \begin{bmatrix} s_1 & \cdots & s_m & \cdots & s_M \end{bmatrix} \in \mathbb{R}^{N \times M} \quad (2)$$

The principal components are then calculated by performing a Singular Value Decomposition (SVD) of the snapshot matrix

$$S = U\Sigma V^* \quad (3)$$

where Σ is a diagonal matrix of singular values and U, V are orthonormal bases of snapshots and

$$U = \begin{bmatrix} u_1 & \cdots & u_M \end{bmatrix}. \quad (4)$$

The optimized special cases are projected to the new base

$$A = \begin{bmatrix} \mathbf{a}^{(1)} & \cdots & \mathbf{a}^{(R)} \end{bmatrix}, \mathbf{a}^{(r)} = \begin{bmatrix} a_1^{(r)} \\ \vdots \\ a_R^{(r)} \end{bmatrix} \quad (5)$$

with the first two principal components shown in Fig. 4 along with the associated optimal algorithm parameters. It can be observed that parameter distribution is non-linear, hence we propose to approximate it by assuming a low-dimensional manifold \mathcal{M}

$$\mathcal{M}(\mathbf{a}) = 0, \mathcal{M} \in \mathbb{R}^R, \quad (6)$$

approximated as

$$\Psi(\mathbf{a}) = p^T(\mathbf{a})c(\mathbf{a}) \quad (7)$$

where p are polynomial basis functions and c are coefficients minimizing the weighted moving least squares criterion [4, 5, 6, 3]

$$J(\mathbf{a}(a)) = \frac{1}{2} \sum_{\mathbf{a}^{(r)} \in V(\mathbf{a})} w(\mathbf{a}^{(r)}, \mathbf{a}) (p^T(\mathbf{a}^{(r)})c - \mathbf{a}^{(r)})^2 \quad (8)$$

and $V(\mathbf{a})$ is a neighborhood defined by the weighing function

$$w(\mathbf{a}^{(r)}, \mathbf{a}) = \exp\left(-\frac{\|\mathbf{a}^{(r)} - \mathbf{a}\|}{2d^{(r)}}\right) \quad (9)$$

where $d^{(r)}$ is the radius of influence of point $\mathbf{a}^{(r)}$.

To determine optimal parameters of the counting algorithm, we take the voxel group $X \times Y \times Z$, project it to the new base obtained in Eq. 3, and find the nearest point on the manifold of parameters (Fig. 4)

$$\mathbf{a}^* = \text{Argmin}(\text{dist}(\mathcal{M}, \mathbf{a}^{(r)})). \quad (10)$$

The parameters at this point are then used to calibrate the counting algorithm. To further increase the accuracy of the approach, the manifold can be refined by adding new expert-validated cases, guided by the geometry of cases back-projected from the manifold.

4 Results

The algorithm has been applied to four datasets : left and right eye of a 3 dpf (days post fertilization), and the left and right eye of a 5 dpf zebrafish (*Danio rerio*) scanned at a $0.743\mu\text{m}$ resolution. Detailed description of the specimen acquisition, scanning setup, and reconstruction is described in [1].

The cell count before and after calibration is summarized in Tab. 1. Without adaptive calibration, the cell count was lower by at least ten percent, reaching over thirty. The default parameters are selected manually, based on a visual appraisal of the accuracy of the count, using samples like the one shown in Fig. 5a for 2D data. But even for a small 2D sample, the heterogeneity of the data set caused 28 cells to remain undetected, with two false positives identified post optimization. After optimization with the initial set of manually-calibrated eight special cases the count was more accurate, although a thorough validation and further refinement of the method are pending. The cell locations detected so far have been reconstructed in 3D reconstruction shown in Fig. 5b.

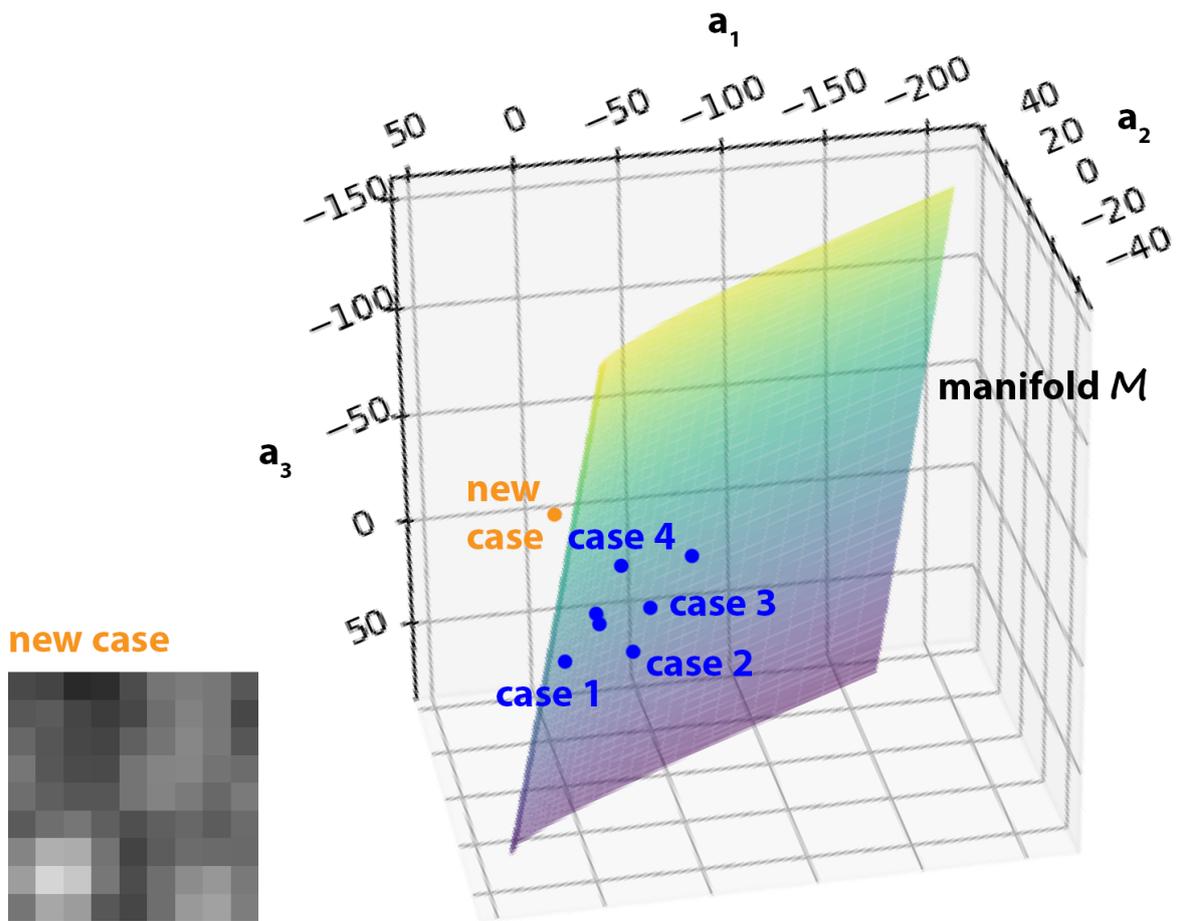


FIGURE 4 – Manifold of the special cases with the new case projected into the reduced-order feature space.

TABLE 1 – Results of the cell count before and after adaptive optimization.

Specimen	Cell count with default parameters	Cell count after optimization	Change
left eye, 3dpf	26 765	30 095	+12.4%
right eye, 3dpf	24 294	29 520	+21.5%
left eye, 5dpf	51 106	68 297	+33.6%
right eye, 5dpf	54 334	63 765	+17.4%

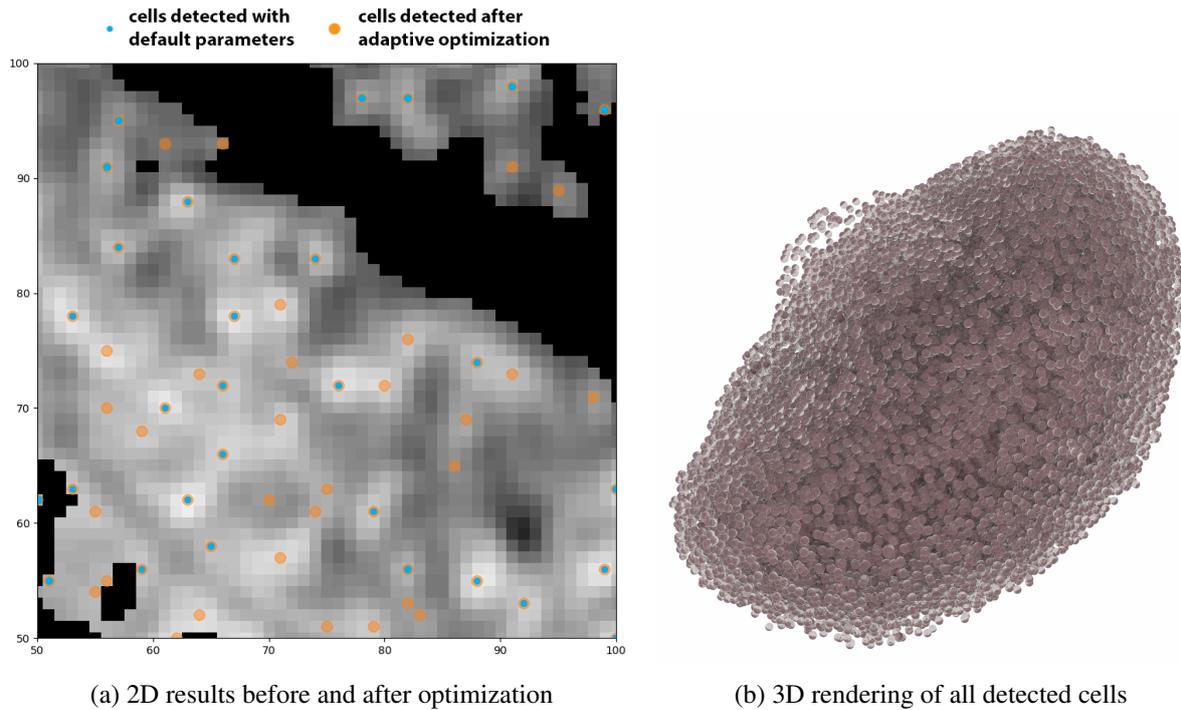


FIGURE 5 – a) Comparison of the 2D results for non- and adaptively optimized cell counting algorithm. b) 3D rendering of cells detected in the left eye of the 5 dpf zebrafish.

5 Conclusions

The proposed approach to calibrate a simple cell counting algorithm has two advantages over more traditional approaches. On the one hand, it manages the bias introduced by the human expert. On the other, it also limits the number of special cases that need to be considered. While it can be argued that this problem can be successfully solved by training a Deep Learning algorithm, our approach empowers non-computational domain experts, allowing them to validate special cases and further refine the solution as necessary. This control over the performance of the algorithm is crucial. It ensures that the quality of the results can be assessed and improved through case validation, a process much more common in medicine and biology, rather than algorithm parameterization and training, ubiquitous in machine learning.

Currently, this methodology for cell counting is being tested on additional cell types in zebrafish specimens. The approach of an algorithm calibrated by a manifold of parameters is universal and could find applications in other fields, from locally adapted X-ray microCT segmentation to local refinement of FEM meshes, everywhere the input of human expert is required to "fix" the algorithm.

Références

- [1] Y. Ding, DJ. Vanselow, MA. Yakovlev, SR. Katz, AY. Lin, DP. Clark, P. Vargas, X. Xin, JE. Copper, VA. Canfield, KC. Ang, Y. Wang, X. Xiao, F. De Carlo, DB. van Rossum, P. La Riviere, KC. Cheng. *Computational 3D histological phenotyping of whole zebrafish by X-ray histotomography*, eLife, 8 :e44898, 2019.
- [2] SR. Katz, MA. Yakovlev, DJ. Vanselow, Y. Ding, AY. Lin, DY. Parkinson, Y. Wang, VA. Canfield, KC. Ang, KC. Cheng. *Whole-organism 3D quantitative characterization of zebrafish melanin by silver deposition micro-CT*, eLife, 10 :e68920, 2021.
- [3] A. Madra, P. Breitkopf, B. Raghavan, F. Trochu. *Diffuse manifold learning of the geometry of woven reinforcements in composites*, Comptes Rendus Mécanique, 346 : 532-538, 2018.
- [4] B. Nayroles, G. Touzot, P. Villon. *Generalizing the finite element method : diffuse approximation and diffuse elements*, Computational Mechanics, 10 : 307-318, 1992.
- [5] P. Breitkopf, A. Rassineux, G. Touzot, P. Villon. *Explicit form and efficient computation of MLS shape functions and their derivatives*, International Journal of Numerical Methods in Engineering, 48(3) : 451-466, 2000.

- [6] B. Raghavan, L. Xia, P. Breitkopf, A. Rassineux, P. Villon. *Towards simultaneous reduction of both input and output spaces for interactive simulation-based structural design*, Computational Methods in Applied Mechanical Engineering, 265 :174–185, 2013.
- [7] S. Rosenhain, ZA. Magnuska, GG. Yamoah, W. Al Rawashdeh, F. Kiessling, F. Gremse. *Data Descriptor : A preclinical micro-computed tomography database including 3D whole body organ segmentations*, Nature Scientific Data 5 :180294, 2018.
- [8] T. Scherr, K. Löffler, M. Böhland, R. Mikut. *Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy*, PLoS ONE 15(12) : e0243219, 2020.
- [9] C. Edlund, TR. Jackson, N. Khalid, N. Bevan, T. Dale, A. Dengel, S. Ahmed, J. Trygg, R. Sjögren. *LIVECell-A large-scale dataset for label-free live cell segmentation*, Nature Methods 18 : 1038-1045, 2021.
- [10] C. Stringer, T. Wang, M. Michaelos, M. Pachitariu. *Cellpose : a generalist algorithm for cellular segmentation*, Nature Methods 18 : 100-106, 2021.
- [11] S. Dimopoulos, CE. Mayer, F. Rudolf, J. Stelling. *Accurate cell segmentation in microscopy images using membrane patterns*, Bioinformatics 30(18) : 2644-2651, 2014.
- [12] QD. Vu, S. Graham, T. Kurc, MN. Nhat To, M. Shaban, T. Qaiser, NA. Koohbanani, SA. Khurram, J. Kalpathy-Cramer, T. Zhao, R. Gupta, JT. Kwak, N. Rajpoot, J. Saltz, K. Farahani. *Methods for Segmentation and Classification of Digital Microscopy Tissue Images*, Frontiers in Bioengineering and Biotechnology 7 :53, 2019.
- [13] F. Lux, P. Matula. *Cell Segmentation by Combining Marker-Controlled Watershed and Deep Learning*, arXiv, 2004.1607v1, 2004.
- [14] AY. Lin, GK. Thomas, KC. Ang, DB. van Rossum, VA. Canfield, KC. Cheng. *Time- and cell-dependent atypia and cell death are caused by progressive deficiency in DNA replication*, bioRxiv, doi : 10.1101/2020.12.31.425001, 2020.
- [15] <https://www.syglass.io> Accessed on Dec. 8th, 2021.
- [16] T. Schwann. *Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants*, Sydenham Society, 1839, translated in 1847.